CLAIMS

What is claimed is:

1.   A system that facilitates analyzing newsgroup clusters, comprising:

a data reception component that receives and recognizes data relating to a plurality of newsgroups; and

an engine that constructs a weighted graph with a subset of the newsgroups represented as vertices of the graph, and cross-postings relating to the subset of newsgroups represented as edges.

2.   A search engine comprising the system of claim 1.

3.   The system of claim 1, further comprising a segmenting component that segments the weighted graph *via* spectral clustering.

4.   The system of claim 3, the segmenting performed as a function of a number of cross-postings between newsgroups.

5.   The system of claim 4, the segmenting component partitioning vertices of the weighted graph into segments so that a total number of edges between different segments is substantially minimized.

6.   The system of claim 5, wherein the segmenting component partitions segments recursively.

7.   The system of claim 3, further comprising a post-processing component that merges a first cluster into a second cluster if a sum of weights between the clusters is greater than a threshold.

8.   The system of claim 7, the threshold being a function of sum of weights of an edge adjacent to the first cluster.

9.   The system of claim 8, wherein two clusters are merged when sum of the weights of edges between a first cluster and a second cluster is more than half of a sum of weights of edges adjacent to the first cluster.

10.   The system of claim 1, further comprising a filtering component that facilitates excluding particular newsgroups from being represented in the weighted graph so as to facilitate reducing the size of the graph.

11.   The system of claim 10, wherein the filtering component excludes newsgroups which do not contain a threshold number of postings.

12.   The system of claim 10, wherein the filtering component excludes newsgroups by utilizing an implicitly trained classifier that infers the type of newsgroup desired by a user.

13.   The system of claim 1, further comprising a paring component that trims edges of the weighted graph with weight less than a threshold weight.

14.   The system of claim 13, wherein the threshold weight is an increasing function of size of the data to be graphed.

15.   The system of claim 14, the paring component removes vertices when the vertices are not interconnected by edges to a threshold number of vertices.

16.   The system of claim 1, upon generation of the weighted graph such weighted graph is relayed to a data store.

17.   The system of claim 16, newsgroup data received by the data reception component is relayed to the data store.

18.   They system of claim 1 outputs the weighted graph to a display device.

19.   The system of claim 18 displays the weighted graph textually.

20.   The system of claim 1, embodied in a computer readable medium.

21.   A method for creating a weighted newsgroup graph comprising:

receiving and recognizing data relating to a plurality of newsgroups; and

constructing a weighted graph such that newsgroups are represented as vertices and cross-posts are represented as edges.

22.   The method of claim 21, further comprising excluding one or more newsgroups from the weighted graph when the one or more newsgroups does not contain a threshold of postings.

23.   The method of claim 21, further comprising excluding one or more newsgroups from the weighted graph by utilizing implicitly trained classifiers.

24.   The method of claim 21, further comprising segmenting the weighted graph into clusters.

25.   The method of claim 24, wherein a spectral clustering algorithm is utilized to segment the weighted graph into clusters.

26.   The method of claim 25, wherein the spectral clustering algorithm is applied recursively to the weighted graph.

27.   The method of claim 26, wherein the spectral clustering algorithm comprises:

calculating vector v by solving an equation $Lv = \lambda Dv$, wherein $L = D - A$ is the Laplacian of the adjacency matrix $A = (\alpha_{ij})$, D is a diagonal matrix with $d_{ii} = \sum_j a_{ij}$, and λ is the second smallest eigenvalue of L;

determining maximum and minimum values contained within vector v;

dividing an interval between the maximum and minimum values of v into Q smaller intervals;

locating a smallest $Mcut$ ratio at endpoints of the Q intervals, wherein S and $\overline{S}$ are two segments resulting from a proposed cut, $cut = \sum_{i \in S, j \in \overline{S}} \alpha_{ij}$, $W_S = \sum_{i, j \in S} \alpha_{ij}$, and

$$Mcut = \frac{cut}{W_S} + \frac{cut}{W_{\overline{S}}};$$

calculating a minimum $Mcut$ ratio of an integer P eigenvector entries before and after the endpoint found to have a lowest $Mcut$ ratio of the Q intervals;

comparing the minimum $Mcut$ ratio of the P eigenvector entries to a threshold t; and

segmenting the eigenvector entry where the minimum $Mcut$ ratio is found if the $Mcut$ ratio is less than the threshold t.

28.   The method of claim 24, further comprising merging the segmented clusters if the weights of edges between clusters is greater than a threshold.

29.   The method of claim 28, the threshold being a function of sum of weights of an edge adjacent to the first cluster.

30.   A system that facilitates analyzing newsgroup clusters, comprising:

a data reception component that receives data relating to a plurality of newsgroups;

an engine that constructs a weighted graph with a subset of the newsgroups represented as vertices of the graph, and cross-postings relating to the subset of newsgroups represented as edges; and further comprising at least one of the following components:

a filtering component that facilitates excluding particular newsgroups from being represented in the graph so as to facilitate reducing the size of the graph;

a paring component that trims edges of the graph with weight less than a threshold weight so as to facilitate reducing the size of the graph;

a segmenting component that segments the graph *via* spectral clustering; and

a post-processing component that merges a first cluster into a segment cluster if a sum of weights between the clusters is greater than a threshold.

31.   The system of claim 30, further comprising a data store for storing at least one of the following:

newsgroup data received by the data reception component;

algorithms utilized for segmenting the weighted graph;

the weighted graph generated by the graphing engine; and

the segmented graph upon the weighted graph being segmented *via* the segmenting component.

32.   The system of claim 30, the post-processing component outputting the modified weighted graph.

33.   A search engine, comprising the system of claim 30

34.   A newsgroup browser comprising the system of claim 30.

35.   An email program comprising the system of claim 30.

36.   A search engine employing the system of claim 30.

37.   A newsgroup browser employing the system of claim 30.

38.   An email program employing the system of claim 30.

39.   The system of claim 30 utilized to facilitate clustering of newsgroups related to buying and selling of goods and services.

40.   A method for creating a cluster graph comprising the following steps:

receiving newsgroup data;

excluding newsgroups that do not contain a threshold number of postings;

paring edges with weight below a threshold;

generating a weighted graph with the newsgroups represented as vertices and the cross-postings represented as edges;

segmenting the graph into clusters;

merging clusters if the sum of the weights between clusters is greater than a threshold; and

outputting the graph.

41.   A system that facilitates analyzing newsgroup clusters, comprising:

means for receiving and recognizing data relating to a plurality of newsgroups; and

means for constructing a weighted graph with a subset of the newsgroups represented as vertices of the graph, and cross-postings relating to the subset of newsgroups represented as edges.

42.   A data packet that passes between at least two computer processes, comprising:

a field that stores a weighted graph representative of a plurality of newsgroups

with a subset of the newsgroups represented as vertices of the graph, and cross-postings

relating to the subset of newsgroups represented as edges